上机4：Python数值分析
1. 开电脑
2. 打开PyCharm

授课老师：

胡晓敏（xmhu@ieee.org）

龚怡（2167570874@qq.com）

明俊峰（34940530@qq.com）

课程资料下载

- 课件

  https://pan.baidu.com/s/1FuQaIoLmT1OnL1hAoBlkVQ

  密码： she6

- Pycharm（或在Jetbrains官网下载，选择Community版本）

  https://pan.baidu.com/s/1iXhXryPJG-YNYF-

  密码

上机4目录

- 理论教学：Python最小二乘法曲线拟合
- 实验：石油产量预测

4.1 拟合的基本过程

- 1. 画散点图

假定某天的气温变化记录如下，使用最小二乘法找出这一天的气温变化规律。

时刻	0	1	2	3	4	5	6	7	8	9	10	11	12
温度	15	14	14	14	14	15	16	18	20	22	23	25	28
时刻	13	14	15	16	17	18	19	20	21	22	23	24	
温度	31	32	31	29	27	25	24	22	20	18	17	16	

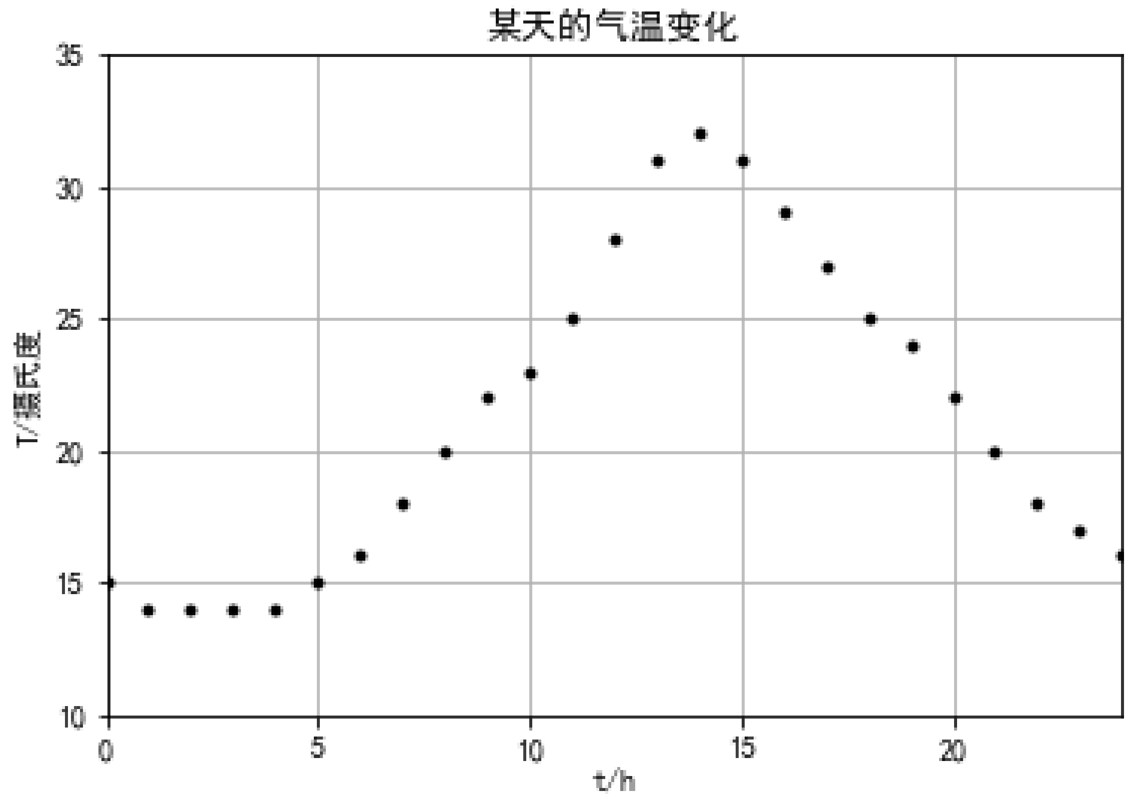
4.1 拟合的基本过程

- 1. 画散点图

```
import matplotlib.pyplot as plt
import matplotlib as mpl
import numpy as np
from scipy.optimize import leastsq    使用最小二乘法对数据进行拟合的库函数

mpl.rcParams['font.sans-serif'] = ['SimHei']  # 指定默认字体
mpl.rcParams['axes.unicode_minus'] = False  # 解决保存图像是负号'-'显示为方块的问题

x = np.arange(0,25,1,dtype='float')
y = np.array([15,14,14,14,14,15,16,18,20,22,23,25,
     28,31,32,31,29,27,25,24,22,20,18,17,16], dtype='float')

# 作图
plt.figure()
plt.title(u'某天的气温变化')
plt.xlabel(u't/h')
plt.ylabel(u'T/摄氏度')
# 坐标轴的范围xmin, xmax, ymin, ymax
plt.axis([0, 24, 10, 35])
plt.grid(True)
plt.plot(x, y, 'k.')
```

4.1 拟合的基本过程

- 1. 画散点图



观察采用何种拟合曲线。

4. 1 拟合的基本过程

- 2. 选择合适的拟合函数类型

二次多项式

$$y = k_1 x^2 + k_2 x + b$$

三次多项式

$$y = k_1 x^3 + k_2 x^2 + k_3 x + b$$

四次多项式

$$y = k_1 x^4 + k_2 x^3 + k_3 x^2 + k_4 x + b$$

自定义函数

$$y = a\exp(-b(x-c))$$

```
# param:最小化的目标| *_fun:拟合函数
# 二次
param0 = [0, 0, 0]
def quadratic_fun(s, x):      # s代表待定系数列表
    k1, k2, b = s
    return k1 * x **2 + k2 * x + b
# 三次
param1 = [0, 0, 0, 0]
def cubic_fun(s, x):
    k1, k2, k3, b = s
    return k1 * x ** 3 + k2 * x**2 + k3 * x + b
# 四次
param2 = [0, 0, 0, 0, 0]
def fpower_fun(s, x):
    k1, k2, k3, k4, b = s
    return k1 * x ** 4 + k2 * x**3 + k3 * x**2 + k4 * x + b
# 自定义函数
param3 = [0, 0, 0]
def myfuns(s, x):
    a, b, c = s
    return a * np.exp(-b*(x-c))
```

k1, k2, k3, k4, a, b, c
都是待定系数

4.1 拟合的基本过程

- 3. 求出拟合曲线（采用leastsq函数）

假设“逼近”规律的近似函数为 $y = f(x)$，即有

$$y_i^* = f(x_i) \qquad i = 1, 2, \ldots, m$$

它与观测值 y_i 之差 $\delta_i = y_i^* - y_i = f(x_i) - y_i$ 称为残差。

残差大大小可以作为衡量近似函数好坏的标准。按照使残差的平方和 $\sum \delta_i^2$ 最小的规则求得近似函数 $y = f(x)$ 的方法称为最佳平方逼近，也称为曲线拟合的最小二乘法。

4.1 拟合的基本过程

- 3. 求出拟合曲线

Python中的leastsq函数可以求出符合要求的最小残差平方和对应的函数系数。

```
# 求出残差
def dist(a, fun, x, y):
    return fun(a, x) - y

funs = [quadratic_fun, cubic_fun, fpower_fun, myfuns]
params = [param0, param1, param2, param3]
colors = ['blue', 'red', 'black', 'green']
fun_name = ['quadratic_fun', 'cubic_fun', '4power_fun', 'a*exp(-b*(t-c))']

for i, (func, param, color, name) in enumerate(zip(funs, params, colors, fun_name)):
    var = leastsq(dist, param, args=(func, x, y)) #求出残差平方和最小的待定系数值
```

4.1 拟合的基本过程

- 4. 检验拟合效果（画图，计算残差）

```
for i, (func, param, color, name) in enumerate(zip(funs, params, colors, fun_name)):
    var = leastsq(dist, param, args=(func, x, y)) #求出残差平方和最小的待定系数值
    plt.plot(x, func(var[0], x), color)
    print('[%s] 二范数: %.4f, abs(bias): %.4f, bias-std: %.4f' % (name,
          ((y-func(var[0], x))**2).sum(),  #二范数的平方
          (y-func(var[0], x)).std(),       #残差的标准差
          (abs(y-func(var[0], x))).mean()) #残差绝对值的均值
    )
plt.legend(['sample data', 'quadratic_fun', 'cubic_fun', '4power_fun', 'a*exp(-b*(t-c))'], loc='upper left')
plt.show()
```

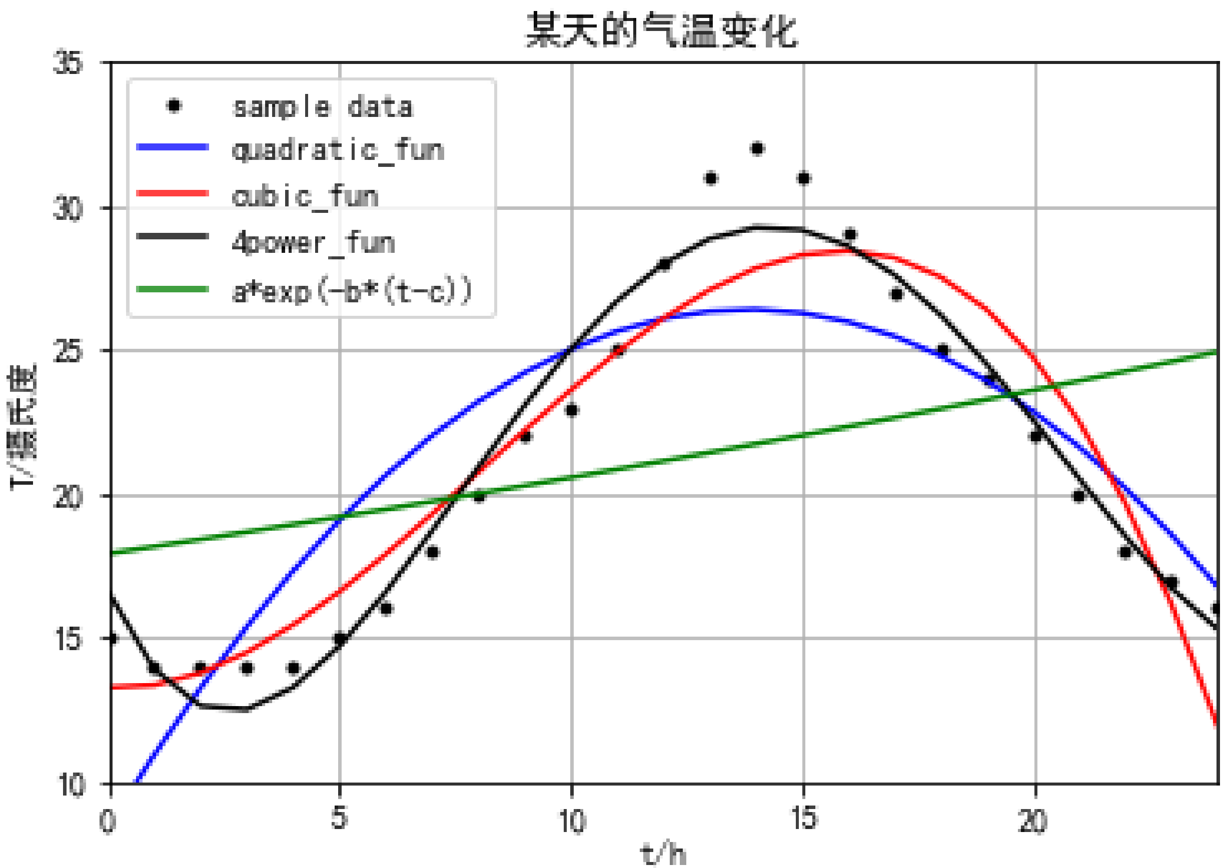
表1　二范数的平方、残差的标准差和绝对值的均值

多项式的类型	二范数的平方	残差的标准差	残差绝对值的均值
二次多项式	241.2443	3.1064	2.5813
三次多项式	106.0781	2.0599	1.6838
四次多项式	36.2837	1.2047	0.9851
T = a * exp(-b * (t-c))	733.6779	5.4320	5.4320

从数值上看，四次多项式的拟合效果最好。

4.1 拟合的基本过程

- 4. 检验拟合效果（画图，计算残差）



4.1 拟合的基本过程

- 实际上，第3步求出拟合曲线，在不使用leastsq函数的情况下，可以根据理论课的方法，对于多项式拟合的情况，计算出正规方程组，并采用直接法或迭代法解这个方程组。
- 对于自定义函数 $y = a\exp(-b(x-c))$，也可以进行线性化转换，类似理论课5.3.2节介绍的

$y = e^{c_0 + c_1 x}$	两边取对数可得 $\ln y = c_0 + c_1 x$	$\ln y = c_0 + c_1 x$

上机4目录

- 理论教学：Python最小二乘法曲线拟合
- 实验：石油产量预测

实验：石油产量预测

- **实验目的：**

比较不同拟合函数的预测精度。

- **实验背景：**

世界石油产量以每天百万桶计，如下表所示，求最佳最小二乘法数值估计：

年	桶/天($\times 10^6$)	年	桶/天($\times 10^6$)
1994	67.052	1999	72.063
1995	68.008	2000	74.669
1996	69.803	2001	74.487
1997	72.024	2002	74.065
1998	73.400	2003	76.777

实验：石油产量预测

- **实验要求：**

1. 分别分析采用(a)直线，(b)抛物线，(c)立方曲线拟合10个数据点的结果，写出拟合后的具体函数表达式。

2. 进行如表1的残差分析。就残差分析结果而言，哪一种拟合最好的代表了这些数据?

3. 利用上面的每一种拟合来估计2010年的石油生产水平，讨论结果。

实验：石油产量预测

思考：本题采用插值方法合适吗？

不合适。石油产量和温度都是带有误差的。

上机课任务

- 提醒：代码需增加详细注释；曲线图注意修改横纵坐标名
- 完成实验内容，整理添加到课程报告内。

并通过email发送给老师

word报告重命名为“学号姓名课程报告(含上机4).docx”

Email标题为“学号姓名课程报告(含上机4)”